Statistical modelling and deconvolution of yield meter data.

FREDE AAKMANN TØGERSEN
Danish Institute of Agricultural Sciences
RASMUS WAAGEPETERSEN
Aalborg University

Abstract

This paper considers the problem of mapping spatial variation of yield in a field using data from a yield monitoring system on a combine harvester. The unobserved yield is assumed to be a Gaussian random field and the yield monitoring system data is modelled as a convolution of the yield and an impulse response function. This results in an unusual spatial covariance structure (depending on the driving pattern of the combine harvester) for the yield monitoring system data. Parameters of the impulse response function and the spatial covariance function of the yield are estimated using maximum likelihood methods. The fitted model is assessed using certain empirical directional covariograms and the yield is finally predicted using the inferred statistical model.

Key words: deconvolution, Gaussian random field, precision farming, prediction, process convolutions, spatial statistics, yield monitoring system.

Running head: Statistical modelling of yield meter data.

1 Introduction

Precision farming is an agricultural practice where e.g. fertilizers and pesticides are allocated to the field according to locally determined requirements. Precision farming brings up a lot of new and interesting applications of spatial statistics since it is required to map many types of variables like yield, weed occurrence, soil properties, root zone capacity etc. In this paper we consider the problem of modelling and mapping the spatial variation of yield in a field. Modern combine harvesters carry a differential global positioning system and a yield monitoring system. (A yield monitoring system is also known as a yield meter — in the sequel we use this shorter term.) These two systems provide corresponding position and yield data. However, certain delay and smoothing effects in the combine harvester must be taken into account when devising a procedure for yield mapping based on yield meter data.

The delay and smoothing effects in the combine harvester are due to the transport time of the grain from the cutterbar (where the crop is cut) to the yield meter and mixing of grain harvested at different locations. A detailed modelling of the grain flow is not feasible and a simplifying assumption is that a yield meter observation is given by a convolution of the unobserved yield and a so-called impulse response function (a kernel integrating to 1), see e.g. Lark *et al.* (1997). Empirical estimation of the impulse response function can be performed by marking crop in a suitably small region with a dye or water and use this as an "impulse" input to the combine harvester. Subsequently, corresponding observations of the proportion of marked yield in the output of the combine harvester and of distance from the impulse region provides an approximation to the impulse response function. Non-parametric estimation of the impulse response function based on this approach is considered in Whelan & McBratney (1997) and parametric estimation in Lark *et al.* (1997), Pringle *et al.* (1999), and Whelan & McBratney (2002).

Attempts to deconvolve yield meter data using an estimated impulse response function and Fourier methods have been made with varying success in Lark *et al.* (1997), Pringle *et al.* (1999), and Whelan & McBratney (2000). The deconvolution problem is ill-posed and Whelan & McBratney (2000) apply presmoothing of the yield meter data in order to reduce the noise sensitivity of the solution. Regularization methods (see e.g. the review in O'Sullivan, 1986) are not considered. After deconvolution a next step would be to create a yield map by interpolating the deconvolved values.

The approach in this paper differs from the previous research in that we consider parameter estimation and deconvolution within the framework of spatial statistics (Ripley, 1981; Cressie, 1993; Stein, 1999). The yield is assumed to be a realization of a random field and a spatial model for the yield meter data is then induced by the convolution of the yield with the impulse response function. Our modelling is thus closely related to recent work in spatial statistics (Higdon, 1998; Higdon *et al.*, 1999; Fuentes & Smith, 2001) where non-stationary random field models are constructed using convolutions of stationary random fields with spatially varying smoothing kernels. We adopt a parametric modelling for both the impulse response function and the covariance function of the yield and compute maximum likelihood estimates of the parameters. Secondly, we use the inferred statistical model to compute the expected value of the yield given the observed yield meter data on a dense grid of locations. This is minimum mean squared error prediction but can also be viewed as a deconvolution where a regularization is imposed by the spatial covariance of the yield. The advantages of our approach are that we obtain parameter estimates without expensive experimental field trials and that the estimated spatial covariance function provides a data driven regularization function.

In Section 2 we describe a specific yield meter dataset and present various exploratory analyses. Modelling of yield meter data is discussed in Section 3. Non-parametric and parametric estimation is considered in Section 4 and Section 5, and deconvolution is discussed in Section 6. The methodology is applied to the yield meter dataset in Section 7. Section 8 finally discusses various open problems.

2 A yield meter dataset

In this paper we consider 3360 yield meter observations from an experimental field at Veddelev belonging to Risø National Laboratory in Denmark. The field was grown with barley and different amounts of nitrogen (0, 54, 90, or 116 kg N/ha) were applied in parallel strips. Each of these strips were harvested separately. The driving tracks of the combine harvester are shown in Figure 1. The 3360 sampling locations on the rather irregularly shaped field are located on the driving tracks with approximately 6 m between each track and approximately 4 m between consecutive sampling locations on the tracks. For convenience the UTM-coordinates have been rotated so that the driving tracks are parallel to the first axis. The empirical treatment means/std. deviations are 3.88/0.51, 5.11/0.46, 5.41/0.42, and 5.54/0.35, for 0, 54, 90, and 116 kg N/ha, respectively. These values indicate that the amount and uniformity of the yield is positively correlated with the amount of nitrogen applied.

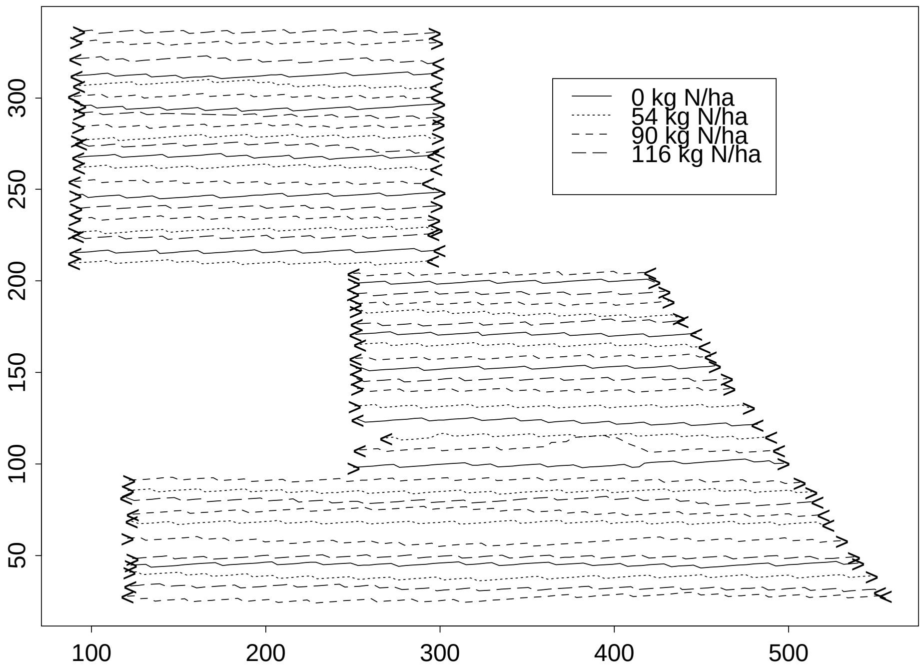

Figure 1: Driving tracks of combine harvester, arrows indicate driving direction.

In order to study the spatial distribution of the yield we first obtain residual yield meter values by subtracting the appropriate empirical treatment means from the observations. Figure 2 shows an interpolated map of residual yield meter values obtained using kernel smoothing with a subjectively chosen bandwidth equal to 15 m. The interpolated values range between -1.41 and 0.93. There is no obvious pattern of non-stationarity in the residual yield meter values.

The univariate distribution of the residual yield meter values appears to be close to a zero-mean normal distribution, see Figure 3. In Section 4 we further consider certain

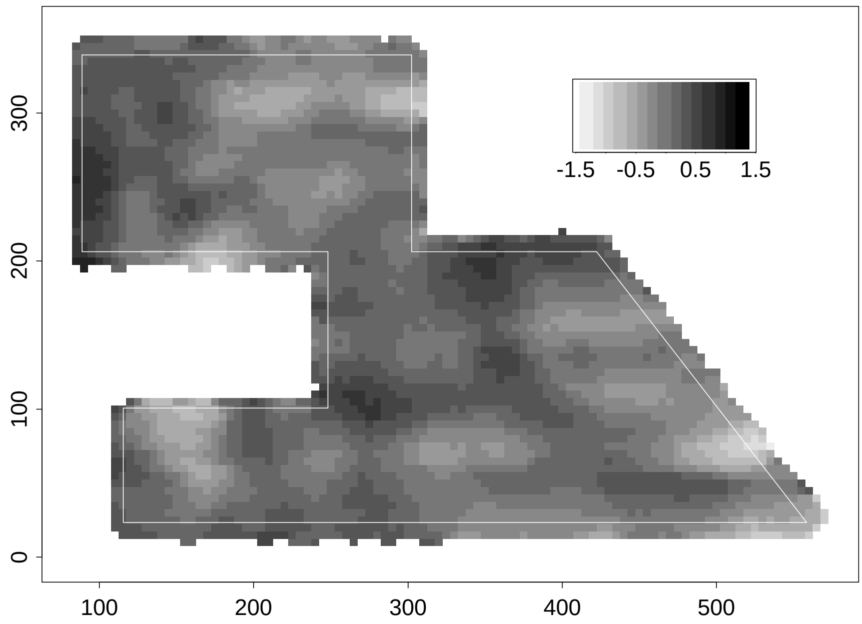

Figure 2: Interpolated residual yield meter values on a 5 $m \times 5$ m grid. Light gray scales correspond to small values, see also the legend in the plot. The white line shows the outline of the field (i.e. the region occupied by the driving tracks in Figure 1).

empirical directional covariograms for the yield meter data.

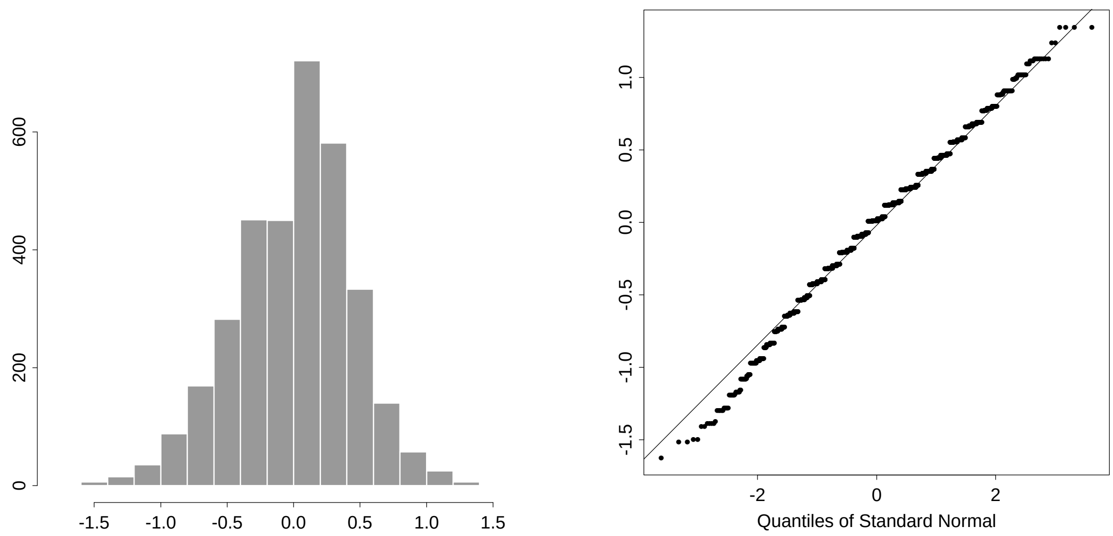

Figure 3: Histogram and quantile plot of residual yield values.

3 Spatial modelling of yield meter data

Following the discussion in Section 1 we assume that the unobserved yield is a realization of a random field $Z = \{Z(x) : x \in \mathcal{D}\}$ where $\mathcal{D}$ is an index set of the field. For a location $x \in \mathcal{D}$ we further assume that the yield meter value at x is given by

$$Y(x) = \int_{\mathcal{D}} \kappa_{d(x)}(x - s)Z(s)ds + \varepsilon(x)$$

where $\kappa_{d(x)}$ is the impulse response function, $d(x)$ denotes the driving direction at x ($d(x) =$ WE, from west to east, or $d(x) =$ EW, from east to west), and $\varepsilon(x)$ is measurement noise. Parametric modelling of the impulse response function is considered in Section 3.1. We refer in the following to $Y = \{Y(x) : x \in \mathcal{D}\}$ as the yield meter process.

For the yield process Z we let $m(x) = \mathtt{E}\big(Z(x)\big)$ and assume that the residual yield process $\delta(x) = Z(x) - m(x)$, $x \in \mathcal{D}$, forms a Gaussian field with constant variance and isotropic covariance function

$$c(\|x - \tilde{x}\|) = \mathtt{E}\left(\delta(x)\delta(\tilde{x})\right), \; x, \tilde{x} \in \mathcal{D}.$$

The residual process δ models variation of the yield caused, for example, by soil properties not accounted for in the mean structure of Z. The parametric model for c is discussed in Section 3.2. The measurement error process $\varepsilon = \{\varepsilon(x) : x \in \mathcal{D}\}$ is assumed to be zero-mean Gaussian with variance σ_ε^2, independent of Z, and with $\mathtt{E}\left(\varepsilon(x)\varepsilon(\tilde{x})\right) = 0$ for $x \neq \tilde{x}$, $x, \tilde{x} \in \mathcal{D}$. The Gaussian assumptions for Z and ε are convenient for computational reasons and do not seem unreasonable when considering Figure 3. The model assumptions are further discussed in Section 7.3. In order to avoid edge effects we consider in the following an infinitely large field $\mathcal{D} = \mathbb{R}^2$.

3.1 Modelling of the impulse response function

The impulse response function κ_d is assumed to be of the product form

$$\kappa_d((z_1, z_2)) = \kappa_{1d}(z_1)\kappa_2(z_2), \; d = \text{WE}, \text{EW}, \tag{1}$$

where $\kappa_{1\text{EW}}(z_1) = \kappa_{1\text{WE}}(-z_1)$, and $\kappa_{1d}(\cdot)$ and $\kappa_2(\cdot)$ are probability densities.

Considering first κ_{1d}, the delay/smoothing effect in the combine harvester may as suggested in Whelan & McBratney (2002) be modelled by a convection-dispersion equation known from mass flow studies in hydrology and soil sciences (see e.g. Jury *et al.*, 1991). The resulting impulse response function is given by an inverse Gaussian $IG(\mu, \lambda)$ density, i.e.

$$\kappa_{1\text{WE}}(z_1) = \sqrt{\frac{\lambda}{2\pi z_1^3}} \exp\left(-\frac{\lambda}{2z_1\mu^2}(z_1 - \mu)^2\right)\mathbf{1}_{\{z_1>0\}} \tag{2}$$

where $\mu, \lambda > 0$ and $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function, see Figure 4 for some plots of (2).

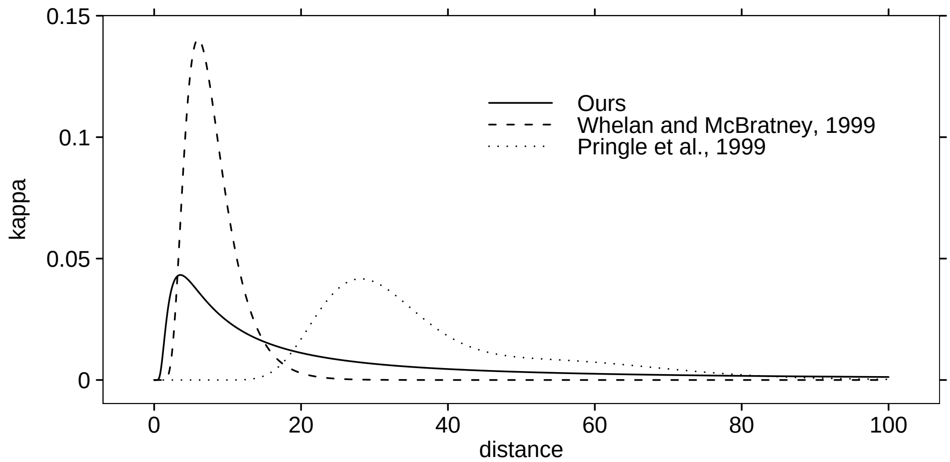

Figure 4: Plot of (2) with $(\lambda, \mu) = (0.64, 8.0)$ (dashed line) and $(\lambda, \mu) = (3.27, \infty)$ (solid line). The dotted line is an impulse response function obtained in Pringle *et al.* (1999) as a mixture of inverse Gaussian densities. More comments are given in Section 7.1.

For simplicity we let $\kappa_2(z_2) = \frac{1}{b}\mathbf{1}_{\{|z_2|<b/2\}}$, i.e. a uniform density where $b > 0$ is the width of the cutterbar. More general functions modelling the smoothing across the driving direction could be chosen in order to capture different delay times across the swath; see Whelan & McBratney (1997) for a study of grain flow in relation to crop position at the cutterbar.

In relation to parameter estimation it is of interest to note that $IG(\mu, \lambda)$ tends to an inverted Gamma distribution with shape parameter $1/2$ and scale parameter $\lambda/2$ when μ tends to infinity (Johnson *et al.*, 1994). This means in practice that if the data dictates high values of μ, then we can effectively only identify λ. It is then advantageous to replace the $IG(\mu, \lambda)$ density with its limiting inverted Gamma density depending just on λ.

3.2 Modelling of the yield process covariance function

The covariance function c of the residual yield process δ is modelled as

$$c(r; \alpha, \tau) = \sum_{l=1}^{L} \tau_l^2 \rho(r/\alpha_l),\ L \geq 1, r > 0, \tag{3}$$

where the components in $\alpha = (\alpha_1, \ldots, \alpha_L)$ and $\tau^2 = (\tau_1^2, \ldots, \tau_L^2)$ are positive and ρ is a known correlation function. Specifically we use a so-called Gaussian correlation function $\rho(r) = \exp(-r^2)$ since this leads to substantial computational simplifications, see Example 1. In order to obtain an identifiable parametrization we order the pairs (τ_l^2, α_l) according to increasing value of α_l so that $\alpha_1 < \alpha_2 < \cdots < \alpha_L$.

The model (3) corresponds to a decomposition $\delta = \sum_{l=1}^{L} \delta_l$ of the yield into mutually uncorrelated zero mean random fields δ_l each with covariance function $\tau_l^2 \rho(r/\alpha_l)$ and exhibiting variation of varying frequency depending on the α_l's. In precision farming especially the slower varying components with large α_l's are of interest since there is a practical limit as to how precisely e.g. fertilizer application can be adapted to high frequency variation in the yield.

3.3 Mean and covariance function of the yield meter process

The mean function of Y is given by

$$\tilde{m}(x) = \int_{\mathbb{R}^2} \kappa_{d(x)}(x-s) m(s) ds, \ x \in \mathbb{R}^2.$$

The covariance between yield meter values at two locations x and $\tilde{x}$ depends on $x - \tilde{x}$ and the directions $d(x)$ and $d(\tilde{x})$ in which the combine harvester was driving when it passed the locations. If $d(x) = d(\tilde{x})$ then $\texttt{Cov}(Y(x), Y(\tilde{x})) = K_0(x-\tilde{x}) + \sigma_\varepsilon^2 \mathbf{1}_{\{x=\tilde{x}\}}$ where K_0 is given by

$$K_0(h) = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \kappa_{\mathrm{WE}}(u) \kappa_{\mathrm{WE}}(v) c(\|h-u+v\|) du dv, \ h \in \mathbb{R}^2, \tag{4}$$

and if $d(x) = \mathrm{WE}$ and $d(\tilde{x}) = \mathrm{EW}$ then $\texttt{Cov}(Y(x), Y(\tilde{x})) = K_1(x-\tilde{x}) + \sigma_\varepsilon^2 \mathbf{1}_{\{x=\tilde{x}\}}$ with

$$K_1(h) = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \kappa_{\mathrm{WE}}(u) \kappa_{\mathrm{WE}}(v) c(\|h-u-v\|) du dv. \tag{5}$$

If $d(x) = \mathrm{EW}$ and $d(\tilde{x}) = \mathrm{WE}$ then $\texttt{Cov}(Y(x), Y(\tilde{x})) = K_1(\tilde{x} - x) + \sigma_\varepsilon^2 \mathbf{1}_{\{x=\tilde{x}\}}$.

Remark 1 The mean function will typically be modelled as a linear regression $m(x) = f(x)'\beta$ where $f(x)$ is a $p \times 1$ vector of known functions and β is a $p \times 1$ parameter vector. Then also $\tilde{m}(x)$ will be given by a linear regression $\tilde{m}(x) = \tilde{f}(x)'\beta$ where $\tilde{f}(x) = \int_{\mathbb{R}^2} \kappa_{d(x)}(x-s) f(s) ds$. In Section 7 f is constant within the driving tracks whereby $\tilde{f} = f$.

Remark 2 The function K_0 is symmetric, i.e. $K_0(h) = K_0(-h)$, $h \in \mathbb{R}^2$. Since δ is assumed to be isotropic, it follows from $\kappa_{\mathrm{WE}}((z_1, z_2)) = \kappa_{\mathrm{WE}}((z_1, -z_2))$ that $K_0((h_1, h_2)) = K_0((h_1, -h_2))$. Combining this with symmetry of K_0 we also get $K_0((h_1, h_2)) = K_0((-h_1, h_2))$. For K_1 we have: $K_1((h_1, h_2)) = K_1((h_1, -h_2))$. These symmetry properties are used for non-parametric estimation of K_0 and K_1 in Section 4.

Using spectral theory for spatial processes (4) and (5) can be simplified to two-dimensional integrals, see Appendix A. For the so-called Gaussian covariance function, further simplification is possible as noted in the following example.

Example 1 (Sum of Gaussian Covariance functions) Under the model (3), both (4) and (5) split into sums of L terms each as (4) or (5) but with c replaced by Gaussian covariance functions $\tau_l^2 \exp(-r^2/\alpha_l^2)$, $r \geq 0$. In Appendix B it is demonstrated that these terms can be reduced to products of one-dimensional integrals where one of the factors is given in closed form. This is very advantageous for computational reasons. Figure 5 shows plots of a Gaussian covariance function $c(r) = \exp(-r^2/15^2)$ and the corresponding functions K_0 and K_1 computed using (21) and (24) in Appendix B.

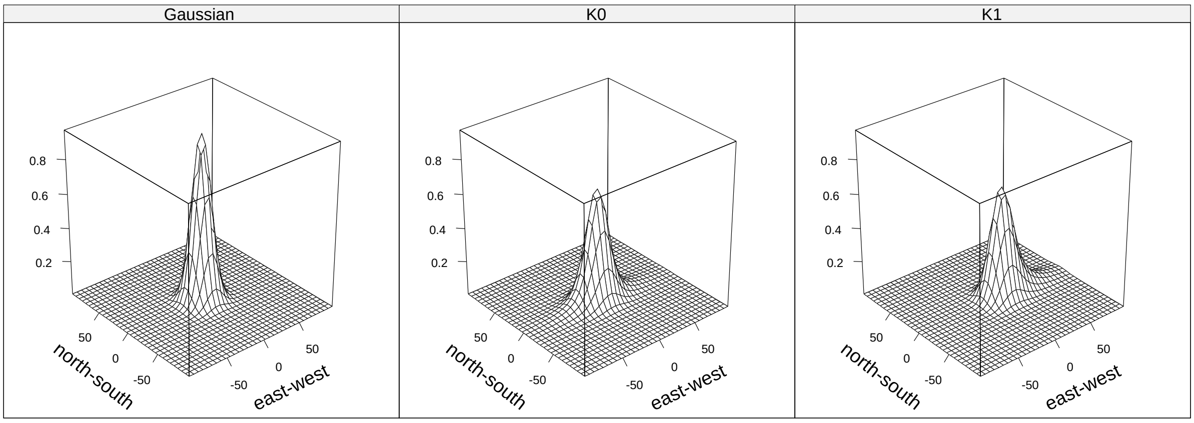

Figure 5: Plots of the Gaussian covariance function $\exp(-r^2/15^2)$, and corresponding functions K_0 and K_1 for $\lambda = 0.05$, $\mu = 10$, and $b = 6$. Note the asymmetry of K_1 in the east-west direction.

3.4 Related work on process convolutions

Modelling spatial processes as convolutions has a long history in spatial statistics with Matérn (1960) (reprinted as Matérn, 1986) as an early reference. Recently, the use of convolutions for constructing models for non-stationary spatial processes has been suggested in e.g. Higdon (1998) and Higdon *et al.* (1999), see also the survey Higdon (2002). Higdon *et al.* (1999) for example consider the following construction

$$\tilde{Y}(x) = \int_{\mathcal{D}} \tilde{\kappa}_s(x)\tilde{Z}(x)ds \tag{6}$$

where $\tilde{Z}$ is a Gaussian white noise and $\tilde{\kappa}_s$ is a bivariate Gaussian density with parameters $\psi(s)$ indexed by s. A prior is imposed on $\{\psi(s) : s \in \mathcal{D}\}$ in order to model a smoothly varying change in $\psi(s)$ across $\mathcal{D}$, and inference is carried out using Markov chain Monte Carlo. A similar construction is used in Hirst *et al.* (2002) in an elaborate space-time model for air pollution in Europe. Fuentes & Smith (2001) consider a model

$$\tilde{Y}(x) = \int_{\mathcal{D}} \tilde{\kappa}(x-s)\tilde{Z}_{i(s)}(x)ds \tag{7}$$

where the smoothing kernel $\tilde{\kappa}$ is not location dependent but where non-stationarity is induced by the index function $i(s)$ which maps s into an index set I of a family of stationary processes $\tilde{Z}_i$, $i \in I$ with different covariance functions c_i.

In Higdon *et al.* (1999) and Fuentes & Smith (2001) the smoothing kernels and fields $\tilde{Z}$ and $\tilde{Z}_i$, $i \in I$, are not of intrinsic interest but merely building blocks for construction of flexible spatial models. This is in contrast to our situation where both Z and κ_d are models of physical realities, i.e. the underlying yield and the smoothing induced by the combine harvester, and where a primary goal is the prediction of Z. Higdon *et al.* (1999) and Fuentes & Smith (2001) use a Bayesian approach for estimation and prediction. This provides a coherent framework for taking into account parameter uncertainty in the prediction variances but for computational reasons the likelihood of the data given the unknown parameters must be easy to evaluate. In Higdon *et al.* (1999) the likelihood is available in closed form while Fuentes & Smith (2001) evaluate the covariance matrix of the data using Monte Carlo integration. In our case the computation of the likelihood is very time-consuming due to the large number of observations. Hence Bayesian inference implemented with Markov chain Monte Carlo is not practical. As in our paper, Hirst *et al.* (2002) compute maximum likelihood estimates by numerical maximization of the likelihood.

In the above discussion we have considered smoothed spatial processes with a continuous index set. In many applications of spatial statistics so-called Gaussian Markov random fields are used partly for computational reasons, see the seminal reference Besag (1974), and Rue (2001) who deals with efficient computation for Gaussian Markov random fields. A Gaussian Markov random field is defined on a discrete grid which could e.g. be a subset of $\mathbb{Z}^2$. In our situation, however, the introduction of a grid seems rather artificial since the yield values are not restricted to a discrete set of locations.

4 Non-parametric estimation of K_0 and K_1

Denote by $\tilde{Y}(x_1), \ldots, \tilde{Y}(x_n)$ approximately zero-mean residual yield meter observations obtained as in Section 2 by subtracting empirical treatment means. Using the symmetry properties of K_0 discussed in Remark 2 we obtain the non-parametric estimate

$$\hat{K}_0(h) = \frac{1}{|N_{0h}|} \sum_{(i,j) \in N_{0h}} \tilde{Y}(x_i)\tilde{Y}(x_j), \quad h \in \mathbb{R}^2$$

where $|N_{0h}|$ denotes the cardinality of the set

$$N_{0h} = \{(i,j) : d(x_i) = d(x_j), \|x_i - x_j - h\| < \phi \text{ or } \|x_i - x_j - (h_1, -h_2)\| < \phi\}$$

and $\phi > 0$ is a smoothing parameter. Note that $\hat{K}_0((0,0))$ is an estimate of $K_0((0,0)) + \sigma_\varepsilon^2$. Similarly for K_1,

$$\hat{K}_1(h) = \frac{1}{|N_{1h}|} \sum_{(i,j) \in N_{1h}} \tilde{Y}(x_i)\tilde{Y}(x_j)$$

where

$$N_{1h} = \{(i,j) : d(x_i) = \mathrm{WE}, d(x_j) = \mathrm{EW}, \|x_i - x_j - h\| < \phi \text{ or } \|x_i - x_j - (h_1, -h_2)\| < \phi\}.$$

The smoothing parameter $\phi > 0$ determines the number of pairs of observations used for the calculation of $\hat{K}_0(h)$ and $\hat{K}_1(h)$.

Figure 6 shows $\hat{K}_0$ and $\hat{K}_1$ computed using the residual yield meter values in Section 2 and with $\phi = 10$ (this value was chosen subjectively in order to get a reasonable degree of smoothness in the estimates). Due to the smoothing done by the combine harvester, $\hat{K}_0$ and $\hat{K}_1$ decrease much slower in the east-west direction (parallel to driving tracks) than in the north-south direction (perpendicular to driving tracks). The asymmetry in the east-west direction in the lower left plot may also be ascribed to the smoothing effect, cf. Figure 5. Note that we plot $\hat{K}_1((h_1, 5))$ and not $\hat{K}_1((h_1, 0))$, $h_1 = -100, -95, \ldots, 100$ m, due to lack of pairs of observations with same second spatial coordinate and, at the same time, different driving directions. For the same reason $\hat{K}_1((0, 0))$ is not plotted. The dotted curves are so-called envelopes used for model assessment in Section 7.3 where we also briefly comment on the bias of the non-parametric estimates.

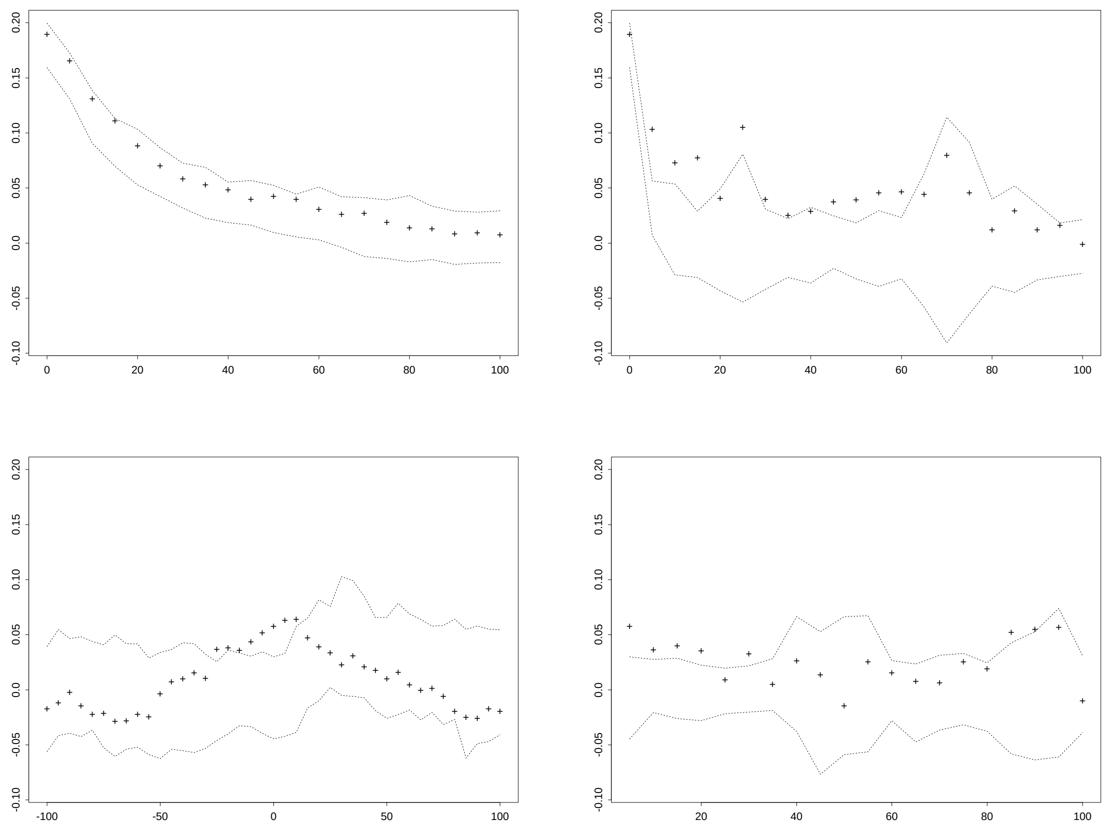

Figure 6: Plots of empirical directional covariograms (+'s) for directions east-west (left plots) and north-south (right plots). Upper plots: $\hat{K}_0((h_1, 0))$ and $\hat{K}_0((0, h_2))$ for $h_1, h_2 = 0, 5, \ldots, 100$ m. Lower left plot: $\hat{K}_1((h_1, 5))$ $h_1 = -100, -95, \ldots, 100$ m. Lower right plot: $\hat{K}_1((0, h_2))$, $h_2 = 5, \ldots, 100$ m. The dotted curves are envelopes used for model assessment in Section 7.3.

5 Parameter estimation

The parameters in the model for yield meter data developed in Section 3 are estimated using maximum likelihood. The sample sizes of yield meter datasets are very large compared with typical datasets in spatial statistics. In the Veddelev data, for example, the sample size is $n = 3360$ and this makes maximum likelihood estimation for the entire dataset impractical. We therefore consider the option of using only a subset of the yield meter data.

For sampling locations $x_1, \dots, x_n$ we let $y = (y_1, \dots, y_n)$ denote the observed realization of the yield meter process $Y^{\text{obs}} = (Y(x_1), \dots, Y(x_n))$. Recalling (3), the covariance function of Y may be written as $\tau_1^2 k$ with

$$k(x, \tilde{x}) = \mathbf{1}_{\{d(x)=d(\tilde{x})\}} k_0(x - \tilde{x}) + \mathbf{1}_{\{d(x)=\text{WE},d(\tilde{x})=\text{EW}\}} k_1(x - \tilde{x}) \\ + \mathbf{1}_{\{d(x)=\text{EW},d(\tilde{x})=\text{WE}\}} k_1(\tilde{x} - x) + \mathbf{1}_{\{x=\tilde{x}\}} \sigma_1^2, \;\; x, \tilde{x} \in \mathbb{R}^2, \quad (8)$$

where $k_0 = K_0/\tau_1^2$, $k_1 = K_1/\tau_1^2$, and $\sigma_1^2 = \sigma_\varepsilon^2/\tau_1^2$. The function k is parametrized by $\tilde{\tau}_l^2 = \tau_l^2/\tau_1^2$, $l = 2, \dots, L$, $\alpha = (\alpha_1, \dots, \alpha_L)$, and the parameters μ, λ of the impulse response function. The covariance matrix of Y^{obs} is denoted $K_{Y^{\text{obs}}}$. The mean function $\tilde{m}(x)$ is modelled by a linear regression $\tilde{m}(x) = \tilde{f}(x)'\beta$, cf. Remark 1.

For a subset $A \subseteq \{x_1, \dots x_n\}$ we let y_A denote the observed realization of the subsampled process $Y_A^{\text{obs}} = (Y(x))_{x \in A}$. Let Σ be the matrix with entries $k(x, \tilde{x})$, $x, \tilde{x} \in A$. The covariance matrix of Y_A^{obs} is then $K_{Y_A^{\text{obs}}} = \tau_1^2(\Sigma + \sigma_1^2 I)$. The maximum likelihood estimates of β, τ_1^2, and $\theta = (\lambda, \mu, \tilde{\tau}_2^2, \dots, \tilde{\tau}_L^2, \alpha, \sigma_1^2)$ maximize the log likelihood function given by

$$l^A(\beta, \tau_1^2, \theta; y_A) = -\frac{n}{2} \log \tau_1^2 - \frac{1}{2} \log |K_{Y_A^{\text{obs}}}| - \frac{1}{2\tau_1^2} \big(y_A - F_A\beta\big)' K_{Y_A^{\text{obs}}}^{-1} \big(y_A - F_A\beta\big) \quad (9)$$

where F_A is the matrix with rows given by $\tilde{f}(x)$, $x \in A$.

For fixed values of θ, the values of β and τ_1^2 that maximize (9) are available in closed form (see e.g. Ripley, 1988). Maximization with respect to θ must be done iteratively. For every new value of θ the entries of the matrix $K_{Y_A^{\text{obs}}}$ must be calculated by numerical integration followed by inversion of $K_{Y_A^{\text{obs}}}$. In an iterative optimization procedure these computations are repeated many times and optimization becomes infeasible if the cardinality of A is large.

Experiments using an approximate likelihood (Vecchia, 1988) based on certain local neighbourhoods were not successful due to problems with choosing appropriate neighbourhood sizes.

In general when data are given by a convolution of an underlying spatially correlated random field one may expect problems with identifiability since a strong correlation in the data may be due to either strong correlation in the underlying field or high degree of smoothing in the convolution. It would then be helpful to collect a moderate sample of direct observations of the underlying field in order to enable identification of correlation parameters and smoothing parameters in the convolution kernel. However, in our case the correlation structure of the yield is isotropic while the highly anisotropic smoothing essentially only takes place in the east-west direction. Pairs of observations in the north-south direction therefore contribute information on the correlation parameters $\alpha_1, \ldots, \alpha_L$ (see also Figure 6). The likelihood function is based on pairs of observations both in the east-west and north-south direction and identification of the correlation parameters $(\alpha_1, \ldots, \alpha_L)$ and the smoothing parameters (μ, λ) is thus possible.

6 Deconvolution

For applications in precision farming it is of interest to map the spatially varying residual yield components δ_l, $l = 1, \ldots, L$, of Z, see Section 3.2. We therefore consider prediction of $\delta_l^* = (\delta_l(u_1), \ldots, \delta_l(u_m))$ where $u_1, \ldots, u_m$ is a grid of locations covering the field. Let $C_{\delta_l^*}$ be the covariance matrix for δ_l^*, and let $K_{Y^{\text{obs}}\delta_l^*}$ be the $n \times m$ matrix of covariances $\mathtt{Cov}\big(Y(x_i), \delta_l(u_j)\big)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. If $d(x_i) = \text{WE}$ then $\mathtt{Cov}(Y(x_i), \delta_l(u_j)) = K_{2l}(x_i - u_j)$ with

$$K_{2l}(h) = \tau_l^2 \int_{\mathbb{R}^2} \kappa_{\text{WE}}(u)\rho(\|h-u\|/\alpha_l)du,\ h \in \mathbb{R}^2, \tag{10}$$

and if $d(x) = \text{EW}$ then $\mathtt{Cov}(Y(x_i), \delta_l(u_j)) = K_{2l}(u_j - x_i)$. The function K_{2l} is computed by factorizing the integral in (10) and applying numerical integration to one of the terms.

The minimum mean square error predictor of δ_l^* is given by the conditional expectation

$$\mathtt{E}\,[\delta_l^*|Y^{\text{obs}} = y] = K'_{Y^{\text{obs}}\delta_l^*} K^{-1}_{Y^{\text{obs}}}(y - \mathtt{E}\,[Y^{\text{obs}}]) \tag{11}$$

and the prediction covariance matrix is given by the conditional covariance

$$C_{\delta_l^*|Y^{\text{obs}}} = C_{\delta_l^*} - K'_{Y^{\text{obs}}\delta_l^*} K^{-1}_{Y^{\text{obs}}} K_{Y^{\text{obs}}\delta_l^*}. \tag{12}$$

The equations (11) and (12) are standard results, see e.g. Mardia *et al.* (1979).

The conditional expectation and variances of δ_l^* are computed in Section 7.4 using Cholesky factorization

$$K_{Y^{\text{obs}}} = L_{Y^{\text{obs}}} L'_{Y^{\text{obs}}} \tag{13}$$

where $L_{Y^{\text{obs}}}$ is lower triangular.

To clarify the connection to regularization note that the conditional mean maximizes the log conditional density of $\delta_l^*|Y^{\text{obs}} = y$. By Bayes theorem the log conditional density $\log f(d_l^*|y)$ is

$$-\frac{1}{2}(y - \mathtt{E}\,[Y^{\text{obs}}|\delta_l^* = d_l^*])' K^{-1}_{Y^{\text{obs}}|\delta_l^*=d_l^*}(y - \mathtt{E}\,[Y^{\text{obs}}|\delta_l^* = d_l^*]) - \frac{1}{2}(d_l^*)' C^{-1}_{\delta_l^*} d_l^* + \text{const} \tag{14}$$

(using obvious notation). In (14) the first term measures fit of the maximizing value of d_l^* and the second term containing the positive definite matrix $C^{-1}_{\delta_l^*}$ acts as a regularization term which ensures a suitable smoothness of the maximizer. Our situation differs from typical regularization problems since the covariance function of δ_l^* plays a role both in the 'regularization term' and in the 'observational model' of $Y^{\text{obs}}|\delta_l^*$. For example,

$$\mathtt{E}\left[Y(x_i)|\delta_l^*\right] = \int_{\mathbb{R}^2} \kappa_{d(x)}(x-s)\mathtt{E}\left[Z(s)|\delta_l^*\right]ds = \int_{\mathbb{R}^2} \kappa_{d(x)}(x-s)\big(m(s) + c_{sl}{}' C_{\delta_l^*}^{-1}\delta_l^*\big)ds,$$

$i = 1, \ldots, n$, where $c_{sl} = \big(\mathtt{Cov}(Z(s), \delta_l(u_j))\big)_j'$.

7 Results for Veddelev yield meter data

For the Veddelev data we consider the model in Section 3.2 with either $L = 1$ or $L = 2$ components in the residual yield process. This reflects a pragmatic approach where first a simple model is fitted and possible deficiencies of the fitted model leads to consideration of a more complex model. In the following we refer to the models with $L = 1$ and $L = 2$ as Model 1 and Model 2, respectively.

The mean function m is assumed to have four levels corresponding to the four nitrogen treatments so that $\tilde{f}(x) = (\tilde{f}_1(x), \ldots, \tilde{f}_4(x))'$ where $\tilde{f}_i(x) = 1$, $i = 1, \ldots, 4$, if the location x received the ith nitrogen treatment and zero otherwise. The width b of the cutterbar of the combine harvester is known and equal to 3.6576 m (12 ft). In the registration of the Veddelev data, observations at both ends of the driving tracks were excluded to avoid edge-effects and we therefore, as in Section 3, ignore the edge-effects. The software for parameter estimation was programmed in Fortran using freely available software for numerical linear algebra (Anderson *et al.*, 1999), optimization (Fox *et al.*, 1978) and integration (Piessens *et al.*, 1983).

7.1 Parameter estimates

We base parameter estimation on observations for a random subsample A of size 840 of the 3360 sampling locations. Both for Model 1 and Model 2 the optimization of the log likelihood moves towards high values of μ whereby an essentially flat likelihood considered as a function of μ is obtained, see Section 3.1. We therefore replace the $IG(\mu, \lambda)$ impulse response function with the inverted Gamma density depending only on λ. In order to assess whether a maximum is reached we inspect profile likelihoods for all pairs of parameters (not shown). The parameter estimates and the values of the maximized log likelihoods multiplied by two are given in Table 1.

Table 1: Maximum likelihood estimates for Models 1 and 2 with respectively one or two components in the residual yield process. Last column contains the log likelihood multiplied by two.

Model	β	λ	τ_1^2	α_1	τ_2^2	α_2	σ_ε^2	2logL
1	(3.89,5.11,5.42,5.55)	4.31	3.40	4.00	-	-	0.008	-334.2
2	(3.76,5.04,5.36,5.47)	3.27	4.03	1.63	0.29	31.74	$4 \cdot 10^{-6}$	-264.9

The likelihood clearly favours Model 2 with two residual components and this is further supported by the model assessment in Section 7.3. The estimate of the correlation parameter α_1 under Model 1 lies between the estimates of α_1 and α_2 for Model 2. For Model 2 the first component δ_1 is practically a white noise with a high variance - this may explain why the estimate of the noise variance σ_ε^2 becomes very small for Model 2 (for numerical reasons we fix σ_1^2 at 10^{-6} whereby σ_ε^2 becomes $4 \cdot 10^{-6}$). The process δ_2 exhibits longer range correlation and has a smaller variance. However, δ_2 exhibits substantial variation compared with the mean function of the yield since the estimated standard deviation τ_2 of δ_2 amounts to 10 % of the highest level of the mean function. The estimated variance of Y is respectively 0.18 and 0.20 for Model 1 and Model 2 and is much smaller than the estimated variance for the yield process Z. This is to be expected since the smoothing in the combine harvester reduces the variance.

A plot of the inverted Gamma impulse response function using the estimate of λ obtained with Model 2 is shown in Figure 4 together with impulse response functions obtained in Pringle *et al.* (1999) and Whelan & McBratney (2002). Pringle *et al.* (1999) considered data from a barley field and their impulse response function is therefore to some extent comparable with ours. The pronounced delay effect for the impulse response function from Pringle *et al.* (1999) does not appear in our case. This is due to an automatic correction for the delay effect built into the data logging equipment of the combine harvester used at the Veddelev field. Whelan & McBratney (2002) considered the crop sorghum, which is very different from barley.

7.2 Parametric bootstrap

To get an impression of the distribution of the parameter estimates under Model 2 we use a parametric bootstrap (Efron & Tibshirani, 1993) as follows: 40 yield meter data sets are simulated from the fitted Model 2 using the Cholesky factorization (13) of the fitted covariance matrix. For each of the simulated datasets, parameter estimates are obtained from a random subsample of size 840 (following the procedure for the original data). We use a moderate number of simulations due to the computational burden of maximizing the log likelihood for each simulated dataset. Table 2 shows medians, means, and standard deviations computed from the 40 sets of bootstrap covariance parameter estimates.

Table 2: Medians, means, and standard deviations computed from 40 sets of bootstrap covariance parameter estimates.

	λ	τ_1^2	α_1	τ_2^2	α_2	σ_ε^2
Median	3.25	3.83	1.63	0.29	31.74	$4 \cdot 10^{-6}$
Mean	3.26	3.97	1.65	0.29	32.56	0.0003
Std. dev.	0.14	0.86	0.18	0.06	3.55	0.0007

The bootstrap parameter means indicate that the parameter estimates are close to being unbiased except for the estimate of σ_ε^2 which has a very skew distribution. All parameters except λ have rather large bootstrap standard deviations relative to the bootstrap means. Concerning the implication of variance parameter uncertainty on prediction results one may note that the prediction (11) only depends on the relative magnitudes τ_l^2/τ_1^2, $l = 1, \ldots, L$, of the variances $\tau_1^2, \ldots, \tau_L^2$. The bootstrap mean/standard deviation for the estimate of $\tilde{\tau}_2^2 = \tau_2^2/\tau_1^2$ is 0.07/0.01. The prediction covariance matrix (12), however, depends directly on τ_l^2.

7.3 Model assessment

The model assessment in this section is based on the non-parametric estimates $\hat{K}_0$ and $\hat{K}_1$ (Section 4) and residuals computed from the full data set with $n = 3360$ observations.

The dotted and dashed curves in Figure 6 are envelopes obtained using a parametric bootstrap under Model 1: the envelopes are point-wise minima and maxima for $\hat{K}_0$ and $\hat{K}_1$ computed from 39 yield meter datasets simulated under the fitted Model 1. If the observed yield meter data were generated under the fitted model, then for each value of h_1 and h_2, there is 95% probability that the pointwise estimate from the observed data falls inside the corresponding envelopes. The envelopes are too narrow since parameter uncertainty is not taken into account. Still, the fit of Model 1 is questionable since in the north-south direction one can observe a large discrepancy between $\hat{K}_0$ and $\hat{K}_1$ computed from the data and their sampling distribution under the fitted model. Figure 7 in contrast shows a good agreement between $\hat{K}_0$ and $\hat{K}_1$ computed from the data and their sampling distribution under the fitted Model 2. In the sequel we exclusively consider Model 2.

Comparison of the theoretical values of K_0 and K_1 with the Monte Carlo mean of the non-parametric estimates computed from 100 simulations of the fitted Model 2 (see solid and dashed curves in Figure 7) indicates a small bias downwards of the non-parametric estimates. This bias is due to that $\tilde{Y}(x_1), \ldots, \tilde{Y}(x_n)$ (see Section 4) are obtained by subtracting the empirical treatment means instead of the true unknown treatment means. The bias is not an issue when $\hat{K}_0$ and $\hat{K}_1$ are used for model assessment.

Residuals are given by $r_i = y_i - \tilde{f}(x_i)'\hat{\beta}$, $i = 1, \ldots, n$, with $\hat{\beta}$ equal to the maximum likelihood estimate. The transformed residuals $(\tilde{r}_1, \ldots, \tilde{r}_n)' = L_{Y^{\text{obs}}}^{-1}(r_1, \ldots, r_n)'$ obtained using the Cholesky factorization (13) should then resemble a sample of n independent standard normal random variables. Figure 8 shows a histogram of the residuals r_i and a quantile-quantile plot of the transformed residuals $\tilde{r}_i$ computed under the fitted Model 2 (in the right plot six transformed residuals with large negative values are omitted). The univariate distribution of the transformed residuals appears to be close to a standard normal.

7.4 Deconvolution

Under the fitted Model 2 the component δ_1 of the yield process is almost a white noise process and exhibits high frequency variation which is not manageable with current precision farming technology. We therefore restrict attention to the slowly varying component δ_2 which is predicted at 6958 locations $u_1, \ldots, u_{6958}$ on a $5\ m \times 5\ m$ grid of dimension 98×71

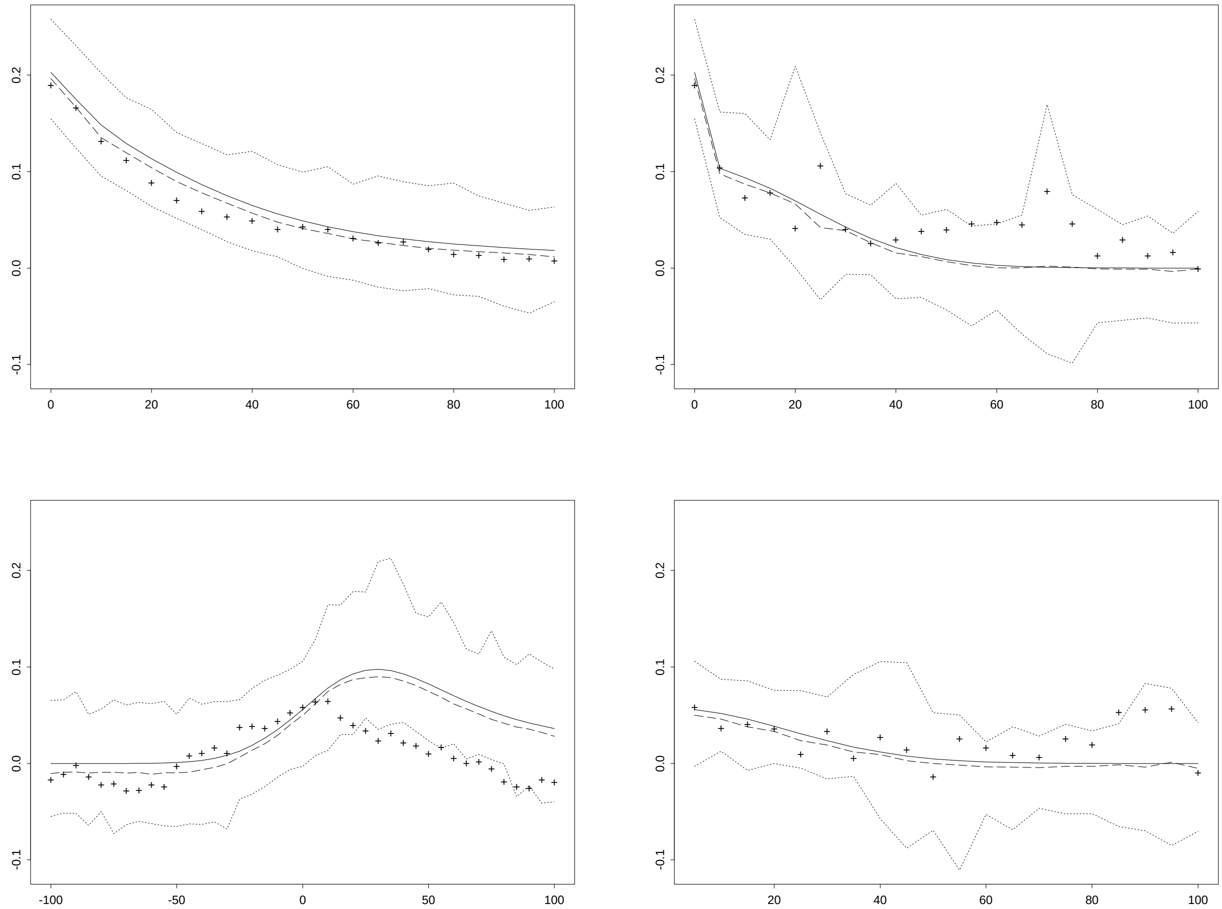

Figure 7: Plot of empirical directional covariograms as in Figure 6 but with envelopes calculated from simulations under the fitted Model 2. The solid line is the theoretical value of K_0 and K_1 under the fitted Model 2. The dashed line is the Monte Carlo mean of the non-parametric estimates computed from 100 simulations.

covering the Veddelev field. Figure 9 shows the prediction of δ_2^* obtained as described in Section 6 and replacing the unknown parameters with the maximum likelihood estimates for Model 2. The predicted values of δ_2 range between -1.47 and 1.36. The variation is thus substantial compared with the fitted mean function of the yield Y and there is scope for using methods of precision farming to e.g. target fertilizer application at areas with low predicted values of δ_2.

Incidentally, the kernel smoothed yield (Figure 2) obtained with a subjectively chosen band width is visually rather similar to the prediction in Figure 9 (note that the kernel smoother used for producing Figure 2 only provides values for locations closer than 15 meter to the field). However, the kernel smoothing does not provide the required deconvolution. The prediction in Figure 9 is furthermore accompanied by prediction variances of $\delta_2(u_j)$,

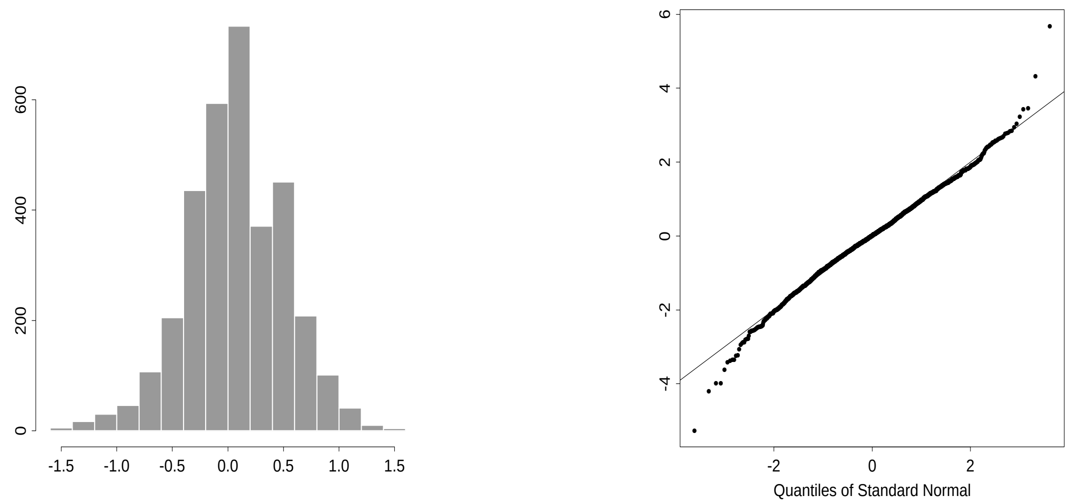

Figure 8: Histogram of residuals r_i (left) and quantile plot of transformed residuals $\tilde{r}_i$ (right, straight line is the identity) computed under fitted Model 2.

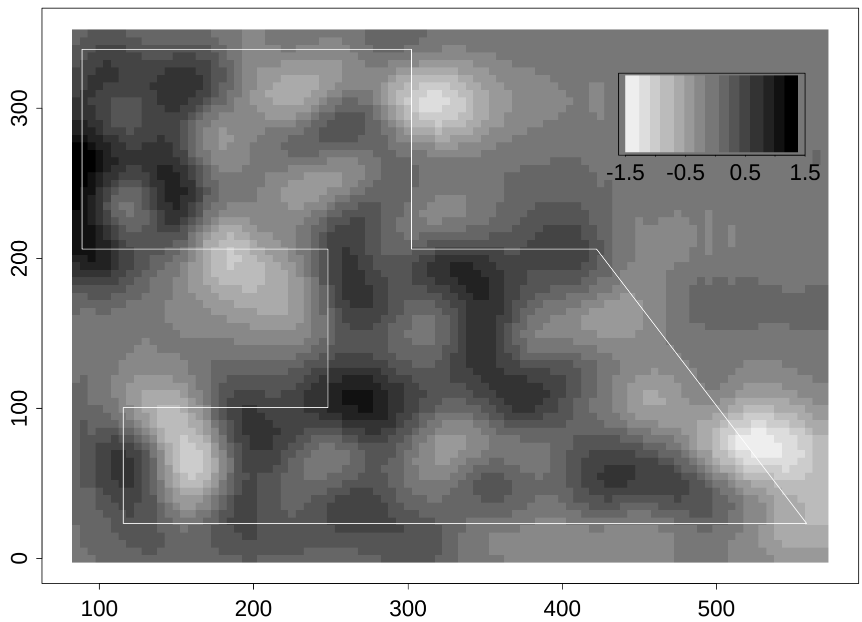

Figure 9: Predicted δ_2^*. Light gray scales corresponds to low predicted values (see legend in plot). White line shows the outline of the field.

$j = 1, \ldots, 6958$ (not shown) which vary from a minimal value of 0.01 up to 0.13.

8 Discussion

For computational reasons we restrict attention to the Gaussian covariance function and a sum of Gaussian covariance functions for which the corresponding random fields are infinitely often mean-square differentiable (Stein, 1999). It is of interest to consider alternatives like spherical or Matérn covariance functions which allow more rough random field realisations. However, for these covariance functions two-dimensional numerical integration is required and this at present leads to unacceptable computing times. Concerning the covariance structure, it could also be relevant to allow for variance heterogeneity between the different treatments.

Edge-effects are ignored in our computations. Considering the heavy tail of the fitted impulse response function this may not be wise. In principle it is straightforward to calculate the marginal distribution for observations near the edges of the field but in practice a lot of extra programming is required. In our analysis we have also ignored the fact that the yield meter observations do not fall on perfectly straight and parallel lines. The deviations from straight lines are mainly due to GPS measurement error so it might be helpful to rectify the data by regressing the sampling locations onto straight lines.

The implications of using a particular parametric form of the impulse response function deserves further investigation. A too heavy tail dictated by the parametric model may e.g. result in an amplification of the estimated yield variance and it would be interesting to consider also an impulse response function with bounded support.

We have focused on using the developed statistical model for yield mapping. Another application of great interest is statistical analysis of experimental field trials where data are obtained using a yield meter. For the Veddelev field for example, the design with the long treatment plots is not used to ensure homogeneous soil properties within plots but is chosen to facilitate data collection using the combine harvester. Since the yield meter data has a non-standard covariance structure, classical analysis of variance does not seem appropriate. It would be interesting to consider alternative statistical procedures based on our approach of modelling the yield meter data.

Acknowledgements We are grateful to Rasmus Nyholm Jørgensen at the Risø National Laboratory in Denmark for providing the Veddelev yield meter data and we thank our colleague Jesper Møller for helpful comments on a previous version of the paper. We also thank the associate editor and the referees for many useful comments that has improved the paper substantially. The first author was supported by DINA - the Danish Network for Informatics in Agriculture and the second author by the Danish Natural Science Research Council.

References

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. & Sorensen, D. (1999). *LAPACK user's* *guide*. SIAM, 3rd edition.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B* **36**, 192–236.

Cressie, N. A. C. (1993). *Statistics for spatial data*. John Wiley & Sons, New York, 2nd edition.

Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Fox, P. A., Hall, A. D. & Shryer, N. L. (1978). The PORT mathematical subroutine library. *ACM Trans. Math. Software,* **4**, 104–126.

Fuentes, M. & Smith, R. L. (2001). A new class of nonstationary spatial models. *Mimeo Series 2534*, Institute of Statistics, North Carolina State University.

Gradshteyn, I. S. & Ryzhik, I. M. (1994). *Table of integrals, series, and products*. Academic Press, New York, 5th edition.

Higdon, D. (2002). Space and space-time modeling using process convolutions. In: *Quantitative Methods for Current Environmental Issues* (eds. C. Anderson, V. Barnett, P. C. Chatwin & A. H. El-Shaarawi), Springer Verlag, London, 37–56.

Higdon, D., Swall, J. & Kern, J. (1999). Non-stationary spatial modelling. In: *Bayesian Statistics 6* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith), Oxford University Press, 761–768.

Higdon, D. M. (1998). A process-convolution approach to modeling temperatures in the north Atlantic ocean. *Journal of Environmental and Ecological Statistics* **5**, 173–190.

Hirst, D., Storvik, G. & Syversveen, A.-R. (2002). Non-stationary spatial modelling. *J. R. Statist. Soc. C* **52**, 377–390.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous univariate distributions*. John Wiley & Sons, Inc., New York, 2nd edition.

Jury, W., Gardner, W. & Gardner, W. (1991). *Soil physics*. John Wiley & Sons, London, 5th edition.

Lark, R. M., Stafford, J. V. & Bolam, H. C. (1997). Limitations on the spatial resolution of yield mapping for combinable crops. *J. Agric. Eng. Res.,* **66**, 183–193.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate analysis*. Academic Press, New York.

Matérn, B. (1960). Spatial variation. *Meddelanden från Statens Skogforskningsinstitut 5/49*, Statens Skogforskningsinstitut, Stockholm.

Matérn, B. (1986). *Spatial variation*. Number 36 in Lecture Notes in Statistics, Springer-Verlag, New York.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science,* **1**, 502–527.

Piessens, R., de Doncker-Kapenga, E., Überhuber, C. W. & Kahaner, D. K. (1983). *QUADPACK. A subroutine package for automatic integration*. Springer-Verlag, Berlin.

Pringle, M. J., Whelan, B. M., Adams, M. L., Cook, S. E. & Riethmuller, G. (1999). Yield deconvolution - a wetted grain pulse to estimate the grain flow transfer function? In: *Proceedings of the Fourth International Conference on Precision Agriculture* (eds. P. C. Robert, R. H. Rust & W. E. Larson), ASA/CSSA/SSSA, Madison, Wisconsin, 1177–1184.

Ripley, B. D. (1981). *Spatial statistics*. John Wiley & Sons, New York.

Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge University Press, Cambridge.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *J. R. Statist. Soc. B* **63**, 325–338.

Stein, M. L. (1999). *Interpolation of spatial data. Some theory for kriging*. Springer, New York.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. R. Statist. Soc. B* **50**, 297–312.

Whelan, B. M. & McBratney, A. B. (1997). Sorghum grain flow convolution within a conventional combine harvester. In: *Precision Agriculture '97 1st European Conference on Precision Agriculture* (ed. J. V. Stafford), BIOS Scientific Publishers, Oxford, England, 759–766.

Whelan, B. M. & McBratney, A. B. (2000). An approach to deconvoluting grain-flow within a conventional combine harvester using a parametric transfer function. *Precision Agriculture* **2**, 389–398.

Whelan, B. M. & McBratney, A. B. (2002). A parametric transfer function for grain flow within a conventional combine harvester. *Precision Agriculture* **3**, 123–134.

Rasmus Waagepetersen. Department of Mathematical Sciences, Aalborg University, Fredrik Bajersvej 7G, DK-9220 Aalborg. email: `rw@math.auc.dk`

A Computation of yield meter process covariance function using spectral theory

By Bochners theorem a covariance function c has a representation as a Fourier transform of a finite measure F on $\mathbb{R}^2$. In the following we restrict attention to the case where F has a density Γ so that

$$c(\|h\|) = \int_{\mathbb{R}^2} \exp(i\omega' h)\Gamma(\omega)d\omega, \ h \in \mathbb{R}^2, \tag{15}$$

where $'$ denotes transpose. Note that $\Gamma(\omega) = \int \exp(-i\omega' h)c(\|h\|)dh/(4\pi^2)$, $\omega \in \mathbb{R}^2$, only depends on $\|\omega\|$.

Using the spectral representation (15) and the product form (1) of κ_{WE} we obtain

$$K_0(h) = \int_{\mathbb{R}^2} |\tilde{\kappa}_1(\omega_1)|^2 \operatorname{sinc}^2(\omega_2 b/2)\Gamma(\omega)\exp(i\omega' h)d\omega$$

where $\operatorname{sinc}(x) = \sin(x)/x$ and $\tilde{\kappa}_1$ denotes the Fourier transform of $\kappa_{1\mathrm{WE}}$. Similarly,

$$K_1(h) = \int_{\mathbb{R}^2} \overline{\tilde{\kappa}_1(\omega_1)}^2 \operatorname{sinc}^2(\omega_2 b/2)\Gamma(\omega)\exp(i\omega' h)d\omega \tag{16}$$

where '$^-$' denotes complex conjugation.

The Fourier transform of $\kappa_{1\mathrm{WE}}$, i.e. the characteristic function of an inverse Gaussian $IG(\mu, \lambda)$ distributed variable, is given by $\tilde{\kappa}_1(\omega_1) = \exp\big(\lambda(1-\sqrt{1-2i\omega_1\mu^2/\lambda})/\mu\big)$ (Johnson *et al.*, 1994). Using the identity

$$\sqrt{z_1 + iz_2} = \pm\big[\big((\sqrt{z_1^2+z_2^2}+z_1)/2\big)^{1/2} + iz_2/\big(2z_1 + 2\sqrt{z_1^2+z_2^2}\,\big)^{1/2}\big]$$

straightforward calculations show that

$$|\tilde{\kappa}_1(\omega_1)|^2 = g_0(\omega_1) \text{ and } \overline{\tilde{\kappa}_1(\omega_1)}^2 = g_0(\omega_1)\exp\big(-ib(\omega_1)\big)$$

where g_0, a, and b are given by

$$g_0(\omega_1) = \exp(2\lambda/\mu)\exp\big(-a(\omega_1)\big), \quad a(\omega_1) = 2\lambda\big((1/4 + (\omega_1\mu^2/\lambda)^2)^{1/2} + 1/2\big)^{1/2}/\mu, \tag{17}$$

and

$$b(\omega_1) = 2\lambda \operatorname{sign}(\omega_1)\big((1/4 + (\omega_1\mu^2/\lambda)^2)^{1/2} - 1/2\big)^{1/2}/\mu. \tag{18}$$

Since a and sinc^2 are even and Γ is isotropic we obtain

$$K_0(h) = 4\int_0^\infty \operatorname{sinc}^2(\omega_2 b/2)\cos(\omega_2 h_2)\int_0^\infty \Gamma(\omega)g_0(\omega_1)\cos(\omega_1 h_1)d\omega_1 d\omega_2. \tag{19}$$

Equation (16) can be rewritten as

$$K_1(h) = \int_{\mathbb{R}} \operatorname{sinc}^2(\omega_2 b/2)\cos(\omega_2 h_2)\int_{\mathbb{R}} \Gamma(\omega)g_0(\omega_1)\exp(i\omega_1 h_1 - ib(\omega_1))d\omega_1 d\omega_2.$$

Expanding the integrand of the inner integral using

$$\exp(i\omega_1 h_1 - ib(\omega_1)) = \cos(\omega_1 h_1)\cos\big(b(\omega_1)\big) - i\cos(\omega_1 h_1)\sin\big(b(\omega_1)\big) + \\ i\sin(\omega_1 h_1)\cos\big(b(\omega_1)\big) + \sin(\omega_1 h_1)\sin\big(b(\omega_1)\big)$$

we observe that the imaginary part of that integral vanishes. Further simplification, using one of the addition formulas for the trigonometric functions, shows that the integrand simplifies to $\Gamma(\omega)g_0(\omega_1)\cos\big(\omega_1 h_1 - b(\omega_1)\big)$. Thus, we obtain

$$K_1(h) = 4\int_0^\infty \mathrm{sinc}^2(\omega_2 b/2)\cos(\omega_2 h_2)\int_0^\infty \Gamma(\omega)g_0(\omega_1)\cos\big(\omega_1 h_1 - b(\omega_1)\big)d\omega_1 d\omega_2. \tag{20}$$

B Calculations for Gaussian covariance function

In general the two-dimensional integrals (19) and (20) must be evaluated numerically. However, for a Gaussian covariance function $c(u) = \sigma^2\exp(-u^2/\alpha)$ some simplifications are possible. In this case the spectral density is given by

$$\Gamma(\omega) = (4\pi)^{-1}\sigma^2\alpha^2\exp\big(-\alpha^2(\omega_1^2+\omega_2^2)/4\big),\ \omega \in \mathbb{R}^2.$$

The integrals (19) and (20) can thus be factorized into the products

$$K_j(h) = \frac{\sigma^2\alpha^2}{\pi}\int_0^\infty \mathrm{sinc}^2(\omega_2 b/2)\cos(\omega_2 h_2)\exp(-\alpha^2\omega_2^2/4)d\omega_2 \\ \int_0^\infty g_0(\omega_1)\cos\big(\omega_1 h_1 - \mathbf{1}_{\{j=1\}}b(\omega_1)\big)\exp(-\alpha^2\omega_1^2/4)d\omega_1,\ j=0,1. \tag{21}$$

By noting that $\sin^2(\omega_2 b/2)\cos(\omega_2 h_2) = 1/2\ \cos(\omega_2 h_2) - 1/4\ \cos(\omega_2(b-h_2)) - 1/4\ \cos(\omega_2(b+h))$, we see that the first integral in (21) may be evaluated as a linear combination of integrals of the form

$$4/b^2\int_0^\infty \omega_2^{-2}\cos(\omega_2 z)\exp(-\alpha^2\omega_2^2/4)d\omega_2 \tag{22}$$

where z takes the form h_2, $b-h_2$ or $b+h_2$. To evaluate (22) note that the integral may be differentiated with respect to z by differentiating under the integral sign since the resulting integrand is absolutely integrable. By Gradshteyn & Ryzhik (1994, 3.952.6)

$$\int_0^\infty \omega_2^{-1}\sin(\omega_2 z)\exp(-\alpha^2\omega_2^2/4)d\omega_2 = \frac{\pi}{2}\,\mathrm{erf}\big(z/\alpha\big). \tag{23}$$

The solution to (22) is found by integrating the right hand side of (23) with respect to z. Thus, by Gradshteyn & Ryzhik (1994, 5.41)

$$4/b^2\int_0^\infty \omega_2^{-2}\cos(\omega_2 z)\exp(-\alpha^2\omega_2^2/4)d\omega_2 = -2b^{-2}\big(\pi^{1/2}\alpha\exp(-z^2/\alpha^2) + \pi z\,\mathrm{erf}(z/\alpha)\big) + c_0$$

where $\operatorname{erf}(x) = 2\pi^{-1/2} \int_0^x \exp(-t^2)dt$ is the error function and c_0 is a constant. Collecting the terms, the first integral in (21) is simplified to

$$\int_0^\infty \operatorname{sinc}^2(\omega_2 b/2) \exp\big(-\alpha^2\omega_2^2/4\big) \cos(\omega_2 h_2) d\omega_2 =$$
$$\frac{\pi^{1/2}\alpha}{2b^2}\Big(\exp\big(-(b-h_2)^2/\alpha^2\big) + \exp\big(-(b+h_2)^2/\alpha^2\big) - 2\exp\big(-h_2^2/\alpha^2\big)\Big) +$$
$$\frac{\pi}{2b^2}\Big((b-h_2)\operatorname{erf}\big((b-h_2)/\alpha\big) + (b+h_2)\operatorname{erf}\big((b+h_2)/\alpha\big) - 2h_2\operatorname{erf}\big(h_2/\alpha\big)\Big) \quad (24)$$

since the contributions from the constant c_0 cancel each other. The second integral in (21) can not be simplified because of the complicated form of the integrand.